Clifford Noble
Embry-Riddle Language Institute (ERLI)
Embry-Riddle Aeronautical University
Daytona Beach, Florida

NONNATIVE ENGLISH-SPEAKING PILOTS' WORKLOAD IN L1 AND L2 AIRSPACE

ABSTRACT

The notion that nonnative speaking pilots incur more communicative workload in dynamic flight than native-speaking pilots was investigated. Evidence is cited primarily from Stanlovitch Dornic's laboratory experiments on bilingualism, and Transport Canada's bilingual IFR-communications simulation experiment, designed largely by Paul Stagar. Dornic found that bilinguals incur greater communicative workload in their non-dominant language (L2). Transport Canada suggests otherwise in situations where bilingual air traffic controllers communicate in pilots' first languages (L1). No other objective real world evidence exists. A dual task paradigm and an attention model are used to articulate the analysis and implications of each experiment. The author encourages development of objective language-workload assessment-devices to supplement subjective evaluations and cockpit design decisions for reducing pilot-workload in L2 airspace.

INTRODUCTION

Common sense leads one to believe that nonnative speaking pilots incur greater communicative workload than native speaking pilots do when flight-communications is in the native speaking pilots' first language (L1) and in the nonnative speakers second language (L2). Evidence was collected from two categories of human factors experiments: laboratory experiments on bilingual communication workload and language dominance, and 2) an instrument flight rules (IFR) simulation experiment by Transport Canada (1979). The IFR simulation required bilingual air traffic controllers (ATC) to communicate efficiently and accurately with unilingual and monolingual pilots on simulated and real flight tasks of normal to high levels of task difficulty. The communication workload metric was based on the frequency of ATC's false starts to unilingual and bilingual pilots (initiating communications in the wrong language), accuracy of communications to the pilots, and accuracy in detecting and correcting incorrect pilot read-backs to ATC. An analysis of both types of experiments is attempted using an attention model and the concept of spare capacity. [1]

The Laboratory Experiments

The results of laboratory experiments of Lambert (1955), Preston and Lambert (1969), Hamers and Lambert (1972), and Dornic (1977-1980). inferred that nonnative English speaking pilots incur greater communicative workload in airspace where the medium of communication is in their non-dominant language (L2). These experiments are discussed in depth.

The limitation of these laboratory-based experiments is that they were task-specific and group specific and difficult to replicate in actual flight; thus, hard to generalize to real-world situations. For example, attitude, motivation, mood, place of testing and the subjects' perception of realism could contaminate laboratory tests. Dornic cited many examples describing these limitations of laboratory studies (Baker, Holding, and Loeb, 1984; Davis and Davis, 1985; Cameron, Robertson and Zacs, 1972). Despite these limitations, the notion that nonnative speaking pilots incur greater communication workload in L2 airspace was supported--indirectly. The discussion now turns to Transport Canada's IFR simulation test.

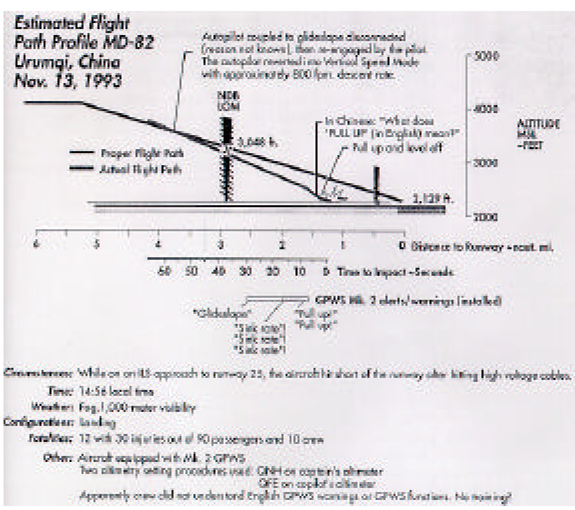
The IFR Simulation Experiment

The Canadian Ministry of Transport's Bilingual IFR Communication Simulation Studies suggested otherwise (1979). Bilingual French-English and English-French controllers were required to speak in L1 or L2 to unilingual and bilingual French and English-speaking pilots. Communication workload appeared the same for the bilingual controllers in both L1 and in L2. Communication workload appeared the same for the pilots who received instructions in their preferred language (L1). False starts and aircraft separation errors were used as a communication workload metric for bilingual controllers. This data was considered relevant to the notion under scrutiny because bilingual air traffic controllers and bilingual pilots collaborate in similar communicative workspace. It was noted, however, that pilot-workload and air-traffic-controller-workload were subject to different variables, and this was a limitation of the study.

The external validity of Canada's IFR simulation experiment was limited because the bilingual-air-traffic-controller subjects were limited to French-

English or English-French speakers. Pilots were limited to French or English speakers. No consideration was given for other Indo-European-based language subjects such as German or Swedish, as in Dornic's laboratory experiments. Moreover, there were no non-Indo-European-based language subjects in either of the experimental categories.

Non-Indo-European language speakers such as Altaic or Sino language-based families comprise Japanese, Chinese, and Korean (Cheng, 1995). They form a growing percentage of world aviation professionals (Dornic, 1980). This posed a safety concern. For example, consider flight in English-only airspace for Japanese pilots whose L2 is English; native English-speaking pilots flying in Chinese airspace who struggle to understand Chinese air traffic controllers' English as a second language (controller L2). Consider a language-related fatality of bilingual Chinese pilots on a localizer approach on 13 November 1993. The Chinese pilots communicated in L1 to native Chinese-speaking air traffic controllers in Urumqi, China. The Chinese pilots had a difficult time understanding unexpected English in Chinese airspace when an automated voice warning systems in English announced: "Pull up, pull up"! (Proctor, 1996). The Chinese pilot's asking in L1, "What does pull up mean?" indicated a lack of knowledge or misunderstanding of intended meaning, but meaning was confirmed too late. How could this have been prevented? Was it a language problem, a training problem or a cockpit design problem? Perhaps a properly designed visual warning would have been better than an L2 auditory warning when a manual task was required (Robinson and Eberts, 1987). Figure 1 shows the details of the accident.



<u>Figure 1.</u> Fatal misunderstanding of automated cockpit warning in English to bilingual Chinese pilots to "Pull-up" while shooting an Instrument Landing System (ILS) approach.

<u>Note.</u> From "Aviation English Seen as Safety Boost," by Paul Proctor, 1996, <u>Aviation Week and Space Technology, 145,</u> p.144. Reprinted with permission.

<u>Laboratory and Simulation Experiments</u>

Neither the laboratory studies nor the simulation study on bilingual workload replicated the real world 100 percent. Although the simulation study was much closer to real world conditions, it was similar to laboratory studies from the standpoint of subject selection being group specific--comprised mainly of Indo-European Ss; thus, not representative of the broader population of non-Indo-European bilingual pilots. Because pilot and controller errors are often the cause of increased workload where errors lead to unplanned or non-aviation specific (non-routine) collaboration, it is likely that Non-Indo-European bilinguals may suffer more communicative workload in Indo-European airspace. Transport Canada (1979) did not focus on this variable of non-routine collaboration for non-Indo-European groups, and this needed to be kept in mind when reviewing its findings. What is non-routine communication?

<u>COMMUNICATION PROBLEMS</u>

<u>Information versus Understanding</u>

Aviation English communication problems exist for native and nonnative English speaking pilots and they cover a wide range of complex considerations (FAA, 1995; Noble, 1997; Ragan, 1997). This is because the language process is inherently complex. For example, information problems that a pilot may encounter in flight may entail situations as complex as a pilot's incorrect perception of a <u>mental navigational model</u>, to a simple inability to hear information. Furthermore, the pilot may misinterpret what he hears due to an incorrect mental navigational model and vice versa. For example, a pilot may think he is north of a station instead of south of a station due to a weakness in mental rotation skills or map reading skills. When ATC refers to the pilot's position as <u>to</u> the station instead of <u>from</u> the station, the pilot may perceive that the air traffic controller made a mistake, and non-routine collaboration would ensue to clarify this information problem. This happens to the best of native speaking pilots in L1 airspace.

<u>Emergencies</u>

It is not surprising that such problems are compounded for unexpected communicative tasks. Emergencies or situations where pilots are required to use their weaker language to attend to routine and non-routine tasks increase workload. It is feasible, therefore, that communicative workload may be greater for nonnative pilots who fly in L2 airspace (Morrow, Rodvold and Lee, 1994).

Mitigation

Another type of communication scenario that involves the dynamics of team communication in hierarchical cockpits is mitigation. For example, a junior flight officer may hesitate to tell the senior pilot of a dangerous situation. He or she may speak in an indirect manner to show respect--or lack of respect. This greatly impacts pilot and copilot behavior--depending on mood, personality, gender and other psychological factors. It affects the efficiency of cockpit resource management (Goguen and Linde, 1983). Such factors are typically present in a wide array of information transfer problems, which are outlined by Billings and Cheaney (1981).

Situational Awareness

In addition to the routine and non-routine communication problems occasioned by mitigation, incorrect navigational models, misunderstandings, and information problems, there is the concept of situational awareness (degree of awareness of events and information transpired, transpiring or having potential to transpire within the cockpit and within the dynamic flight environment), which can be reduced by potential communication problems or piloting problems. This is because a pilot's situational awareness is dependent upon his or her efficient and accurate use of language resources and piloting skills--most likely a combination of both (Endsley (1988).

A comprehensive understanding of all of these issues is necessary to gain insight into the magnitude and complexity of communication workload. A discussion of each of these related topics is beyond the scope of this paper. This paper focuses on an in-depth analysis of experiments on bilingual communication workload. The next section elaborates on the terms routine and non-routine collaboration.

Elaboration on Non-routine Collaboration

Non-routine collaboration is undesired and is usually the result of misunderstanding, or information problems in flight. Non-routine communication may cause increased pilot workload, rather than being a symptom of increased workload (Hart and Bortolussi, 1984: 554-556). Morrow et al. (1994: 254) said, "This may be particularly disruptive when controller and pilot have different native languages, because they may resort to different conversational strategies..." leading to a snowball effect (Clark and Schaeffer, 1987). The point is that there are more variables to control in the real world, and if simulation studies claim ability to predict real world outcomes, they should be designed to control all the additional variables--such as those mentioned concerning workload outcomes of unplanned pilot error. Appendix A identifies the type and frequency of pilot and ATC controller communication problems within controlled airspace. Appendix B lists some language related aviation incidents or accidents (Noble, 1997). The next section discusses an overview of the laboratory experiments on bilinguals.

OVERVIEW OF LABORATORY STUDIES

This next section outlines in chronological order laboratory experiments conducted on bilinguals. It starts with a study that demonstrates that language dominance can be measured in bilinguals. It concludes with studies that reveal hidden costs to bilinguals for attaining superior overt performance on L2 tasks. These apparent balanced bilinguals look to be as proficient on L2 tasks as monolinguals are on L1 tasks. The costs are measured by an attention model, which employs the concept of spare capacity and a dual task paradigm.

In between these two discussions on establishing that language dominance can be measured, and measured indirectly with a great degree of accuracy through an attention model, are discussions of experiments that heighten the reader's awareness of linguistic issues; especially those that impact bilingual and monolingual communicative workload. Such considerations usually entail efficiency, which is expressed in terms of response time (RT), and accuracy constraints on language performance. Examples are: (1) the theoretical bilingual switch mechanism, which some scholars argue makes it impossible for a bilingual to process information in a second language while attending to communicative tasks in another language, (2) the sensitivity of bilinguals to semantic similarities in L1 encountered while attending to L2 communicative tasks--and the resulting propensity for increasing RT; (3) the sensitivity of Indo-Europeans bilinguals to similar words in different Indo-European languages, and the costs associated with exposure to those words on a linguistic task--compared to Non-Indo-European speakers' indifference to Indo-European sounds in L2.

The discussion builds upon the attention model as a powerful and practical methodology for controlling for complex psychological considerations while enabling a finer linguistic measure of language ability. One hopes this methodology will be applied in a practical way with the help of technology to predict language ability while reducing risks associated with flight training. The paper now turns to the laboratory experiments, of Lambert (1955), Preston and Lambert (1969), Hamers and Lambert (1972), and an array of experiments by Stanlovitch Dornic, on language dominance, noise, and the capacity to attend to communicative tasks. Stanley Donic's experiments span from 1969 to his death in 1989.

REVIEW OF LABORATORY EXPERIMENTS

Measuring Language Dominance

Lambert (1955) activated a Ranschburg exposure apparatus and a reaction-time clock with a control key to expose 32 random French and English instructions to 52 bilingual subjects. The purpose of the experiment was to measure Ss' L1 and L2 decoding time (comprehension time) of visual instructions. The subjects of Lambert's study were three groups of 14 students each classified by degree of experience in French. Group U was native-English-speaking undergraduate French majors; Group G, native-English-speaking graduate French majors; and Group F, native-French-speakers from Europe with French baccalaureates and a minimum of seven years stay in an English-speaking country. The circular Ranchburg exposure apparatus randomly exposed 16 commands in English and 16 commands in French. Ss responded by key-press. Subjects followed typed instructions in L1 and in L2, after which they pressed one of eight keys defined by position and color. Ss placed their eight fingers on the eight keys (similar to the manner in which a typists place finger on the keyboard of a typewriter) and attended to the instructions from the circular Ranschburg apparatus. Only one key was to be pressed at a time.

Before the Ss began the experiment, they practiced reacting to visually presented instructions that referred to position (left or right hand) and numbers. Numbers were placed on Ss' eight fingers (thumb excluded). The fingers were labeled from each index finger starting with the number one to the small finger, which was labeled four. "Right four" would indicate that the little finger on the right hand should depress the key beneath it, and so forth. As soon as Ss pressed a key in response to what they saw in the circular Ranchburg exposure apparatus, the reaction time was recorded, another random command was exposed, and the clock started timing anew.

For the actual experiment, the keys on the right were labeled by colors of red, green, black and yellow. The keys on the left were labeled by colors of green, red, yellow and black for the left fingers (from index finger to little finger, respectively). Ss were instructed to look for and comply with typed instructions on a card, which the experimenter controlled from the circular Ranchburg device. Instructions might be "left red" in English or "gauche, rouge" in French, in which case Ss needed to decode and respond by pressing the appropriate key. The experimenter did not explain how the accuracy of the Ss' response (which keys were pushed) and response times (RTs) were recorded.

The experimenter used a one-tailed test to obtain t scores because all Ss' background information was obtained prior to selection and participation in the experiment. Two tailed tests were considered inappropriate by the experimenter because the experimenter felt there was sufficient S background information.

Calculations from Table 1 showed that Ss' mean response time (RT) was slower for L2 tasks than for L1 tasks. RT was faster with more experience in the target language (L2). If the t-score was above .05, the subject was considered to be dominant in one language; if below the .05 level, Ss were considered to be balanced bilinguals. The experiment confirmed what Cattell had observed in 1877 of bilingual English-German speakers--that there was a cost associated with the linguistic behavior of second language speakers. Cattell had noted that it took native English speakers who were familiar with German a surprisingly long time to associate a word in English (L1) to its German (L2) equivalent. This cost was best measured in terms of speed--response time. The faster the RT, assuming accuracy, the more automatic the linguistic behavior. The slower the RT, the less automatic the linguistic behavior (Lambert, 1955: 197).

Lambert's experiment allowed for statistical analysis and scoring of language dominance among bilinguals. He used the following formula to account for individual differences in absolute reaction times. :

$$[\ (\sum \text{RT: L2}) - (\sum \text{RT: L1})\ ] \div (\sum \text{RT: L1}) = \text{Ss score}$$

In other words, the difference of the total response times for L1 and L2 was divided by the total time of the faster language (L1) to account for individual differences in absolute reaction times. With this adjustment, the average mean for the difference scores were greater (RT was longer) for the less experienced language (L2) in each group. As more experience was gained in a second language, the reaction time of language tasks became faster.

The implications were threefold. First: response time is a useful tool for measuring performance and workload of bilinguals. Second: for experiments that seek to compare performance of monolinguals and bilinguals in dynamic domains, it would be very important to establish a base-line for language competence in both L-1 and L-2 before generalizing the findings to the population under study. Third: the method of determining what constitutes a "balanced-bilingual" must be carefully considered in the analysis of each experiment and when comparing findings among similar experiments.

The Language Switch Mechanism

Preston and Lambert (1969) conducted research on the controversial "switch mechanism" in bilinguals and its effect on RT when switching languages. They created a bilingual version of the Stroop color-word task (1935) to examine the functional differences between bilinguals' L1 and L2, and to discover if the activation of L1 would make L2 inoperative, or if the activation of L2 would make L1 inoperative. The original Stroop color-word task was a procedure whereby Ss receive three cards. Ss had to perform tasks A, B, and C whereby:

Task A: read the names of colors-words that were printed in black. Task B: name the colors of batches or blocks of colors. Task C: name the colors of the inks in which color-words were printed.

In task C, for example, the word red would appear in blue on the card, and the S needed to say "blue." Task C was the most difficult because color recognition and reading were two tasks that competed with the Ss' attentional resources.

Preston and Lambert's modified Stroop color-word test used only steps B and C. Step B was for bilingual skill-classification only. No significant differences among the Ss' bilingual skill levels were discovered using Card B, which was used as partial support for presuming that the Ss were balanced bilinguals.[2] The validity of the presumption remains questionable due to the simplicity of the task. The study required three experiments.

The Ss of the first experiment were eight English-French balanced bilinguals and eight English-Hungarian self-proclaimed balanced bilinguals. Three forms of Card C were used where translations of English, French and Hungarian appeared on separate cards. 100 words appeared on each card in 10 rows and 10 columns each. English words were red, blue, green and brown. French words were rouge, bleu, vert and brun. Hungarian words were piros, kek, zold and barna. Each language represented 25 words per card and no two words or colors appeared next to each other. With all these cards, the Ss had to do six tasks: 1) name the colors of English color words using English, 2) name the colors of Hungarian or French color-words using English, 3) name the colors of the asterisks from Card B using English, 4) name the colors of English color-words using Hungarian or French, 5) name the color of French or Hungarian color-words using Hungarian or French and 6) name the colors of asterisks in Hungarian or French.

The six tasks were assigned to the subjects in random order to eliminate order effects. Card B tasks were executed first to help insure balanced bilingual skill level. Card B consisted of three forms where an asterisk for the color of the letters from each of the three C cards was represented. Table 2 shows the mean time scores for the six tasks for English-Hungarian and English-French.

The first experiment showed that the bilinguals' performance on interlingual and intralingual tasks depended on degree of skill in both languages and upon the similarities of stimulus on information presented in either language.[3] The findings showed that balanced bilinguals suffered interlingual interference due to a tendency to translate printed words and to pronounce them. Pearson and Lambert presumed that the bilingual subjects were balanced in L1 and L2 based on two criteria: a) the Ss' self-perception of equivalent reading skill performance in L1 and L2, and b) on Ss' performance on the second task, the medium difficulty task of the three. Self-perception was later found to be a good indicator of skill level (Albert, 1978)[4]; however, the low task difficulty of task 2 challenged the accuracy of Ss' bilingual skill-level upon which the findings of the experiment relied.

Notice that the mean response times for tasks 2 and 4 of the English-Hungarian bilinguals were lower than for tasks 1 and 5. Tasks 2 and 4 response times resulted when different languages were used for stimulus and response between an Indo-European language and a non-Indo-European language. The difference is that since the similarity stimulus of the words were dissimilar, the translation processing time was less because there was less interference caused by similarity of stimulus in each language. The case was the opposite for English-French bilingual stimulus and response because the stimulus words were more similar to the words in English or French, causing more interference in the translation process.

The second experiment used English-German bilinguals to examine the suggestion of the first experiment that the tendency to expend processing time translating lessened when the translation of words in color had stimulus characteristics that were different. Antithetically, Indo-European language-based words that were similar like English blue and French bleu took longer to process because of the interference caused by the similarity of the words and the tendency to translate them. Therefore, the tendency to translate was less for words with less similarity. This was confirmed between bilingual English-German Ss with two sets of C Cards prepared as in experiment one, but with first set color words being similar, such that: English words were green, red, blue and brown. German were grun, rot, blau and braun. In set two, the English words were black, yellow, pink and purple. The German words of the second set were schwarz, gelb, rosa, and lila. As was expected, the words of the first set were so similar that the interference times were roughly the same in all four conditions involving similar words in English and German (conditions 1,2,3 and 4). See Table 3 for details. On the other hand, the words that were more dissimilar in English and German took less time to respond to due to less interference and the resulting decrease in tendency to translate dissimilar words.

The major implication here is that depending on the skill of the bilingual subject, words in one language that sounded like words in another language or words in the same language that were juxtapositions could cause interference and increase response time and error rate. Consider this situation in the flying domain: The instructor points out the window and says "Cotton gin!" The student does not notice the instructor pointing out the window, but hears him say, "Cut engine!" (FAA, 1988). If this stimulus produced a question in the student pilot's mind a non-routine collaborative scheme would transpire (Morrow et al., 1994); thus, increasing reaction time to attend to the instructor's intended meaning--and increase workload potential. Consider another situation where a Japanese pilot hears his instructor say "way down." The sound of the English word "way" is the same pronunciation for the Japanese word "up" (/uei/). So there is potential for translation interference even for balanced bilinguals if the Stroop word-color test can be generalized to verbal sounds in context. In this situation, "way down" would be initially translated to "up down," which would cause some interlingual interference and potential time delay in understanding or complying with navigational tasks that are initiated with verbal input where the pilot's situational awareness were not optimal in difficult airspace conditions.

The third experiment tested the dominant English bilingual English-French subjects' decoding and encoding efficiencies to discover if there were other factors beside stimulus similarity that slowed down reaction time in the bilingual Stroop color-word test. Dominant English subjects demonstrated greater efficiency to decode; thus, there was a higher probability of interference for dominant English subjects to encounter interference on words that were similar to English words on the Stroop color-word test--in reaction time and error rate.

However, when non-color words were used, encoding time appeared to be significant only when tasks occurred between different languages. This finding implied that some process other than encoding translated words was responsible for interference because there were no errors in reading or in translation categories with non-color words across languages. See Table 4.

This suggested that difficult tasks (simulated by color in the Stroop word-color test) enhanced intralingual and interlingual interference, but that simpler tasks may not levy enough difficulty to make a significant difference in linguistic performance in L1 or L2; therefore, when applied to real world tasks in the flight domain, the piloting tasks would apparently need to be of sufficient difficulty for language interference to become a factor (Dornic, 1980b; Morrow et al. 1994). This difficulty level is a function of the bilingual pilot's piloting skill and language skill.

Semantic Interference

Hamers and Lambert (1972) adapted the Stroop color-word test (1935) to a bilingual pitch-word test and demonstrated that balanced bilinguals were unable to ignore the semantic aspects of verbal stimuli and context. Using eight high-and-low-pitched pronunciations of the English words high and low, and the French words haute and basse, they discovered that latency effects (RT) were greater for verbal-semantic characteristics and context-based-semantic characteristics than they were for tones (pitch sounds) alone.

The subjects were 16 French and English balanced bilinguals. Ages ranges from 17 to 25 years of age. Ss were self-proclaimed balanced bilinguals who had similar second language acquisition ages. To add control, Hamers and Preston administered Ss three Basic-category pitch-word tasks for screening out unbalanced bilinguals. They used the two most difficult of three tasks to make their determination. Only Ss who attained mean latency scores (RT) within 100 milliseconds between tasks B2 and B3 were selected. See table 5.

Task B2 required Ss to say "high" or "low" after hearing either English word high or low. Task B2 word-stimuli were accompanied by either a low-pitched voice of 110 cycles per second (cps) or a high pitched voice of 175 cps. Task B3 required Ss to say "basse" or "haute" after hearing either French word basse or haute. Task B3 word-stimuli were accompanied by either a low-pitched voice of 110 cps or a high pitched voice of 175 cps. The Basic-task-category was the simplest category of three: 1) Basic, 2) Control and 3) Interference categories.[5]
Each S was presented 15 tasks from combinations of these eight verbal stimuli. The eight verbal stimuli were 1) the English word high with a pitch of 110 cps, 2) the English word high with a pitch of 175 cps, 3) the English word low with a pitch of 110 cps, 4) the English word low with a pitch of 175 cps, 5) the French word haute with a pitch of 110 cps, 6) the French word haute with a pitch of 175 cps, 7) the French word basse with a pitch of 110 cps and 6) the French word basse with a pitch of 175 cps. The bilingual experimenter pre-recorded all eight and controlled the tape recorder from a separate room.

The 15 tasks were comprised of three Basic tasks (B1, B2 and B3), six Control tasks (C1, C2, C3, C4, C5 and C6), and six Interference tasks (T1, T2, T3, T4, T5, and T6). The three Basic and the six Control tasks were repeated 100 times, and the six Interference tasks were repeated 200 times for each S The Basic, Control and Interference tasks were of increasing difficulty, respectively. Table 5 presented a legend of the tasks and the response modes.

The three controlled response modes were 1) pressing a key, 2) speaking in English and 3) speaking in French. In the Key Press mode, Ss were required to press one of two keys that were on a board when they heard any of the eight stimuli. One key produced a high pitched tone of 800-cps, and the other key produced a low tone of 200-cps. The experimenter did not describe how the Ss distinguished between the two keys--only that Ss were trained to know which key produced a high pitch and which key produced a lower pitch. The Ss were instructed to press the key that corresponded to the pitch of the words that were presented aurally over a tape recorder. The resulting tone produced by Ss pressing either key was recorded on a tape recorder.

The Ss were in a separate room with earphones on their head, a microphone to speak into, and a handheld board with two keys available to push. The Ss' room was wired to record the Ss' responses on a separate response tape, which also recorded the stimuli. Stimuli were presented randomly by the experimenter. The experimenter allowed four seconds between each stimulus, but allowed the Ss more time to respond if a response was not made within four seconds. A timer was activated by a voice-activated relay in the experimenter's room, and a separate voice-activated relay was in the Ss' room for timing Ss' responses (RT). Ss were tested one at a time in two sessions.

The first session was for two hours with the three Basic tasks. The Basic tasks were delivered in the order of B1, B2 and B3; and then B1, B3 and B2 to eliminate order effect. The second session ranged from 1.5 hours to 1.75 hours. English and French stimuli were alternated so half of the subjects received French stimuli first and vice versa. The same applied to the order of the six Control tasks and six Interference tasks.

The results of the experiments were presented in Tables 6 through 9. The results in Table 6 showed the mean latencies for the 15 tasks. A one-way analysis of covariance and the Sheffe multiple comparison procedure confirmed that RT was always longer for responding to verbal stimuli than to non-verbal stimuli (words versus tones).

A two-way analysis of variance was performed on four language levels (all the combinations and orders of English and French language stimulus) and on three stimulus-task levels (congruent and noncongruent task/stimulus combinations). Table 7 showed the results. No significant difference was found in the mean RT of either language, French or English, for this type of task. As a result, the experimenter combined the mean RT results of French and English and showed the mean RT for combined-language-tasks in Table 8, where the data showed that interference (increase in RT) occurred even without Ss' verbal responses in either English or French. In other words, all that was required to cause interference was semantic similarity between the

combinations of verbal stimulus and the semantic similarity associated with pitch. This highlighted that RT increased without language encoding tasks (language production tasks). This was relevant because it was obvious that tasks that involved (decoding + encoding) would take longer than decoding alone because decoding is a prerequisite to encoding (unless one speaks before thinking).
Table 9 showed data to demonstrate that the Ss were balanced bilinguals. There was no significant difference in the response accuracy of either combination of monolingual or bilingual stimulus. The RT, however, was the hidden cost of accuracy. What needed to be reckoned with was that balanced bilinguals had a longer RT for monolingual-stimulus (as opposed to bilingual-stimulus). This accentuated the finding that RT was longer when two stimuli had relevant meaning. Balanced bilinguals who would operate in a unilingual environment would find more of this type of load because more monolingual-stimulus exists in a unilingual environment than in a bilingual environment.

The implication for bilingual pilots in a real world flying-context was clear. First: More attentional resources would be required to attend to any semantic stimuli in non-verbal noise with L1 or L2 than would be required in a context of noise without L1 or L2. There would be greater workload for tasks that required encoding and decoding than for decoding alone, but there would still be a price to pay for partial decoding--even if the pilot never responded.

Second: If the semantic aspects of a verbally initiated stimulus (primary stimulus) prompted a primary task in a context that anticipated continued use of the same primary stimulus, then balanced bilingual pilots would be unable to ignore any secondary verbal-stimulus in L2 (such as that which might be encountered in a flight requiring intra-cockpit or inter-cockpit L1 / L2 communications. This would increase potential for increased workload and for increased RT for a given task. Given the same situational context, the L1 pilot would have less semantic stimulus to attend to, and more spare capacity to attend to anything else.

Hamers and Lambert demonstrated that pure tones alone levied a price on reaction time, but that semantic-sounds cost more in terms of attentional resources needed to attend to a primary task. For bilingual pilots, the issue was whether or not the aspects of semantics of L2-prompted verbal-stimuli demand more attentional resources when attending to a primary task in noise than did the aspects of semantics of L1-prompted-verbal-stimuli in noise. Dornic (1979) demonstrated that there was a higher cost for balanced bilinguals in noise when the semantic stimuli was in a second language (L2) and in simple counting of numbers of flashes, tones and vibrations in L2 (1969). The former will be discussed first.

Language Dominance in Simple Workload Tasks

Dornic (1969) gathered data on bilinguals' dotting error (tendency to underestimate the number count in higher frequencies of stimuli instantiations). Nine self-pronounced balanced-bilingual Ss who had at least six years of study of the English language counted the number of units of three stimuli: flashes, tones and vibration in either L1 or L2 (Swedish or English), as instructed by the experimenter. The experimenter instructed Ss to count covertly. Ss were instructed to refrain from storing the numbers in memory or counting overtly (aloud).

Visual stimuli (white flashes) were presented through a 1-centimeter diameter circular hole that was positioned one meter from the Ss. The verbal stimuli (tones) were presented over a set of earphones. The Ss felt vibrations by placing their left index finger on an eight-millimeter vertical rod. Each stimulus lasted for 11 msec. each time it was applied as part of a 2 second round stimulus of 13, 15 or 17 total stimuli per 2 sec round. Each stimulus was presented three times.

It was the Ss' job to discover the number of stimuli through silent counting in either English or in Swedish, as instructed by the experimenter. The purpose of the experiment was to test the hypothesis that internal pronounceability of stimulus had a marked effect on performance in the areas of perception, learning and memory--and in particular, increasingly so on L2 speakers. The hypothesis was that L2 RT should increase, and Ss' accuracy should decrease as Ss' ability to pronounce words in L2 decreased. Findings reconfirmed the hypothesis of Martin and Schultz, 1963; Schultz, 1965; Martin 1966 and the suggestion of Kovac, 1965 and 1966 (as cited in Dornic, 1966) that coding RT increases as stimuli become more difficult to pronounce covertly in L2. Table 10 showed the arithmetic means of all the Ss' reported counts of each of the three stimuli conditions. Figure 2 simplifies the averages over sense modalities of L1 and L2.

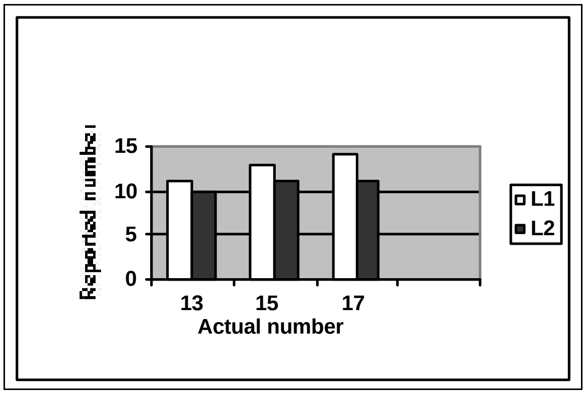


Figure 2. Average vs. actual number of stimuli of flashes, vibrations and auditory senses counted by self-proclaimed balanced English and Swedish bilinguals in a covert counting task.

The implications suggest the dominance of the first language in balanced bilinguals over a second language for simple workload tasks such as counting numbers, flashes, tones and vibrations. All of these phenomena are present in flight, but the important point is not the phenomena itself, but the implication that increased workloads created in a dynamic environment have the potential to enhance the increased RT it would take a bilingual to attend to an unexpected navigational task, emergency or non-routine collaborative schema requiring communication in L2 for either. "The whole coding system is simply slower, less prompt and economical in the nondominant language..." (Dornic, 1969: 397).

Decoding Time and Detection Performance

Dornic (1980a: 121-122)[6] cited a 1978 experiment where he tasked bilingual subjects to detect certain combinations and sequences of digits with verbal stimuli in L1 and in L2 with specific rules. Decoding time and detection performance were worse in L2 even though the subjects had rated themselves to be "balanced bilinguals." The results are demonstrated in the Figure 3.

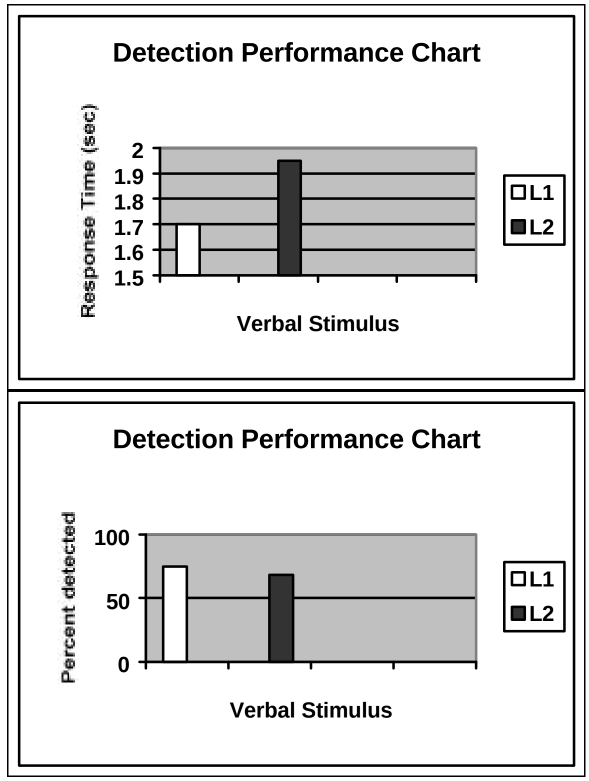


Figure 3. Response time and performance on random L1 and L2 decoding and detection workload tasks with verbal stimulus.

Bilinguals' RT was slower and detection performance of checklist items lower when the verbal stimulus was in the nondominant language. Detection performance was more accurate and RT faster when the verbal stimulus was in bilinguals' dominant language. Dornic noted that this held true for high frequency elementary L2 words that might be expected to be automatic. He attributed the latency in detection performance of elementary L2 words to semantic content, which took longer to decode. Although accuracy and efficiency margins between L1 and L2 verbal tasks appeared insignificant; the accumulation of errors of these simple processes were believed to lead to chain reactions because "basic information processing stages determine the course of more complex mental processes" (Dornic, 1980a: 120).

Effects of Quiet and Noise

Dornic (1978) conducted five experiments on the effects of quiet and noise on bilingual's RT in L1 and L2. In all five experiments Ss were classified as “balanced bilinguals” whose L2 performance was at least 85% of their L1 performance on the same tasks. The subjects were bilingual Indo-European family class members from groups of Swedish-English,

English-German and German-English, German-French and French-German; and a mixture of Indo-European with Slovak based languages; specifically, Czech (Slovak)-German, and Czech (Slovak)-Swedish). Noise was a mixture of conversational pieces, typing, office and street noise. Words could not be deciphered in the noise.

Covert Pronounceability in L1 and L2. The first experiment presented 10-15 unrelated words visually, one at a time, two words a second, in either of the bilingual's L1 or L2 language. The task was aimed at measuring immediate memory. In quiet, task performance was equal in L1 and L2, but L2 performance in noise was significantly lower than L1 performance in noise. The results of the first experiment noted that even though the words were presented visually, L2 performance in noise was significantly inferior to L1 performance in noise. See Table 3 for the statistical methodology.

Dornic's explanation (1969) was that covert pronounceability of words in L2 was less efficient than covert pronunceability in L1 because noise masked subvocal L2 speech acts that would normally enhance performance. This made a strong implication for pilots using their second language. If covert pronounceability were inferior in L2 tasks in noise, then bilingual pilots' L2 tasks in noise would appear to suffer in stressful situations where time sensitive responses to verbal ATC tasks were critical. This would also apply to visual L2 tasks like map reading and responding to cockpit instruments that use L2 scripting. Map reading was considered one of the higher workload tasks among experienced native speaking pilots (Hart and Bortolussi, 1984).

Masking Effect of Noise on Inner Speech. The second experiment was a verbal task in quiet and in noise. Ss rehearsed unrelated words while simultaneously transforming numbers with a set of rules. This created a heavy load. Performance in quiet showed little difference in L1 and L2, but inferior performance was more obvious for L2 in noise. This could be "attributed to the masking effect of noise on inner speech" (Dornic, 1978:3). Consider the process a pilot goes through in pilot-read-backs. Pilots often echo the command by internal speech. If verbal commands are masked by a bilingual pilot's inner speech, or by other noise masking the subordinate language, then bilingual pilot-accuracy and pilot-efficiency might suffer in high workload situations such as emergencies in complex airspace where primary task difficulty is high.

Noise Strengthens L1 Tendency. The third experiment (Dornic, 1978) was an encoding task (speech production). Subjects were presented with linguistic stimuli in the form of pictures, which were considered neutral linguistic stimuli. Ss were required to name the pictures in L1 or L2 when they were presented again. Pictures were presented in-groups of three. Noise was played three minutes before the pictures were presented, and it was played throughout the presentation of the pictures. The result of the Ss' performance in noise made the negligible difference between L1 and L2 in quiet more pronounced. Noise strengthened the tendency to use or rely on the stronger language (Broadbent, 1971). In the case of an in-flight emergency, it stands to reason that a bilingual pilot would tend to think and act in the stronger language. This would decrease pilot efficiency and accuracy and would be a function of language dominance. These experiments were conducted on balanced bilinguals; therefore, nonbalanced bilingual-pilot-performance in L2 airspace could be expected to be significantly lower than balanced-bilingual-pilot performance under the same conditions.

Translation to L1 in Sustained Noise. The fourth experiment was based on Wicken's concept of "release from proactive inhibition" (as cited in Dornic, 1978). The effect of sustained noise on language dominance pronounced the automatic tendency of bilinguals to translate from L2 to L1. In other words, the longer that bilinguals were exposed to noise, the greater the tendency of bilinguals to translate from L2 to L1.

Ss were presented three words that were unrelated. Then the Ss were distracted for 20 seconds in another verbal activity. After the distraction was completed, the subjects were required to recall the unrelated words. This process was repeated four times with different words for each trial. Trials one through three used one of the Ss' languages, and the fourth trial used the other language. The findings showed that recall was best in the fourth trial when the language changed--for both quiet and noise--but recall was worse for the fourth trial when language changed in noise alone. The significance of this finding implied that as bilingual pilots fly and operate in increasingly complex airspace where more noise and stress preside, bilingual pilot communication efficiency and accuracy would decrease.

Switching Time Greater in L2. The fifth experiment was typical of bilingual tasks that involved prompt switching between L1 and L2 where Ss performed a "continuous verbal task" in one of their two languages (Dornic, 1978: 5). Ss needed to stop the continuous verbal task in the one language and name verbally presented stimuli such as numbers, colors and pictures, using the other language. Switching time from the dominant to the subordinate language and from the subordinate to dominant language was made in noise and in quiet. Switching time was greater in noise and in quiet for L2, but slightly greater in noise. This emphasizes the role of bilingual pilots and crewmembers who act as a team in the cockpit. RT for the L2 verbal component instantiated by ATC may take longer in L2 airspace than in L1 airspace. The same concept could be applied to flight in uncontrolled and controlled airspace.

An Analogy. If an analogy could be made to real world flight, quiet could represent the type of airspace in which a bilingual pilot might fly, where there were little or no communicative interaction with air traffic control--such as flight in uncontrolled airspace. Noise could represent the type of airspace in which a bilingual pilot might fly, where moderate to heavy communicative interaction with air traffic control existed...where required pilot-controller interaction were increasingly frequent or demanding...where more frequent aircraft requests, read-backs and expected compliance with time sensitive communication tasks competed over the same frequency. The implication ensued that bilingual pilots and bilingual crews might find themselves at greater risk in unexpected, non-routine situations in L2 airspace before they would in L1 airspace because their demand for limited attentional resources would exceed supply sooner in the former airspace than in the later.

Dornic demonstrated that noise was a major factor that increased bilinguals' reaction time in L2, and this suggested inferior performance in complex, demanding task domains or situations where information input rate was too high. Overload for decoding tasks might slow down a bilingual's time to understand verbal input. The same was true for speech tasks where the searches for verbal labels in L2 slowed down when time pressure was a factor.

It is important to recognize that even though these differences in performance may be small when viewed individually, "they can nevertheless be very important in professions requiring, for example, prompt reaction to verbal input" (Dornic, 1980: 370). Comprehension and speaking tasks are typical of basic in-flight communication tasks for an analysis of the types of information load encountered in flight see Morrow, Rodvalld and Lee (1994).

L1 and L2 Decoding and Detection

Dornic (1980a) measured the overall time it took bilinguals to check items on a list after they had received verbal instructions in their dominant and non-dominant languages. Verbal L1 and L2 instructions identified position, color, shape or value. The response time was slower for all L2 tasks. The simplest task was illustrated in Figure 4 where the subject needed only to comply with instructions manually by selecting the appropriate items.

Left two	
2 4 3 1	1 2 4 3
Left one	
3 4 1 2	2 1 3 4
Right four	
3 2 1 4	4 3 1 2
Left three	
2 1 3 4	1 4 3 2
Right one	
4 1 3 2	2 1 4 3
Right three	
3 1 4 2	4 1 2 3
Left four	
1 3 4 2	2 4 3 1
Right four	
2 3 4 1	2 1 3 4
Left one	
2 1 3 4	3 4 2 1

Figure 4. Response time and performance on random L1 and L2 decoding and detection workload tasks with verbal stimulus.

L2 Motor Response.

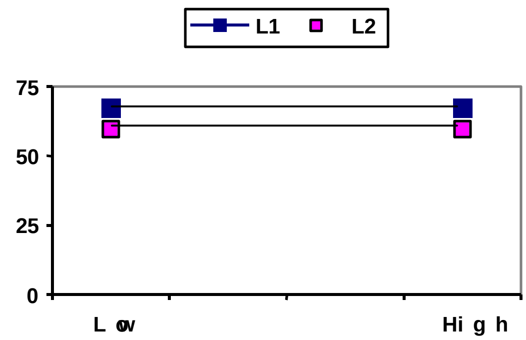
Inferior language performance may affect other tasks--based on the assumption that attentional resources are limited. Dornic (1977) had 44 bilingual subjects press buttons when they were presented unexpected non-verbal aural and visual signals while simultaneously attending to a continuous verbal activity. The continuous verbal activity was either conversational engagement with the experimenter or creating simple language sets in L1 or L2. Motor response took longer when the subjects were engaged in L2. The more complex and difficult the motor response, the greater the difference between L1 and L2; therefore, L2 performance was a function of the task difficulty of the motor skill. This suggested that there was less spare capacity to attend to anything else in L2 than there is available in L1.

Covert Costs and Spare Capacity
Dornic (1980b) conducted an experiment to discover if there were hidden costs associated with balanced-bilinguals' overt performance on tasks where the verbal component was essential. He employed a methodology that proposed to measure the hidden costs of effort (limited attentional resources) using the concept of spare capacity, a time sharing paradigm, and Kahaneman's single-resource-theory (1973). The theory assumed that each person possessed a limited amount of attentional resources beyond which no resources were available to attend to any other task--or to improve performance on a current task--no matter how much additional effort was expended. Attention and effort were synonymous. Spare capacity was the quantity of limited attentional resources remaining after attending to a primary and a secondary task simultaneously--the quantity of remaining limited resources that could be used to attend to anything else.

Ss were 28 self-proclaimed balanced-bilingual Ss. The first independent variable was 60 short typed instructions, which constituted the primary task (PT). The instructions were presented visually to the Ss. The visual instructions were commands that required Ss to check-off items that were described by color, shape, position or value. Ss were asked to complete the task as quickly as possible. The experimenter manipulated the PT in two manners. One control manipulation was the quantity of information presented in the visual instructions. The second manipulation was the type of stimulus-response compatibility-- either L1 or L2.

The second independent variable that was controlled and manipulated by the experimenter was a series of auditory frequency irregularities and clicks of varying intensity. The clicks and frequency changes were played over a tape recorder at 65 d B and were designed so as not to interfere with the input of the primary independent variable or the Ss' responses. Ss were required to attend to the experimenter's manipulations (intensity changes and frequency irregularities) by means of a secondary tapping task (2T). Tapping on a desk was the response mode from which the 2T dependent variable (performance measure) was obtained.

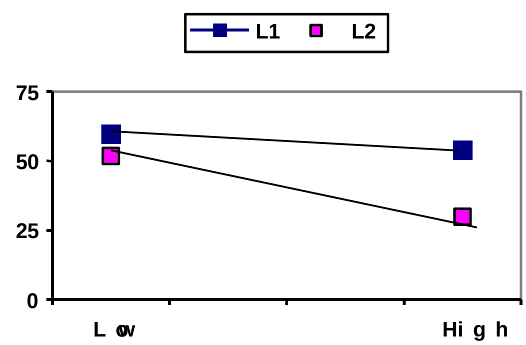
After the exercise was over, Ss were asked to estimate their perceived effort on the primary task (PT), which was the checking off of items defined by color, shape, value or position in the visual instructions. Ss used a psychometric scale of 0 through 10, which represented degree of difficulty from minimal effort to maximum effort, respectively. The Ss' perceived efforts of the PT represented the dependent variable (performance measure) of the PT. The Ss' tapping accuracy and tapping efficiency represented the dependent variable (performance measure) of the 2T. Figures 5 displays PT and 2T performance.



PT Task complexity

Figure 5. Primary task performance at low to high difficulty levels for L1 and L2

Balanced bilinguals, by definition, performed as well as monolinguals on overt tasks as indicated on the PT task complexity chart in Figure 5, but not without covert cost, as indicated on 2T in Figure 6.



2T Task complexity

Figure 6. Secondary task performance at low to high difficulty levels for L1 and L2

The hidden performance costs were insignificant when viewed alone, but when viewed as cumulative, the differences in performance could be very significant i.e., performance in L2 airspace where variables surrounding noise and stress are not controlled. This was considered especially true of complex, unplanned tasks of increasing difficulty such as in emergencies requiring time-sensitive reaction to verbal input, or timely decoding of verbal stimulus from air traffic controllers. The monolingual pilots would incur an additional load as well, but the bilinguals might incur substantially higher workload and a higher performance deterioration rate if the covert costs provided less capacity to cope with the added variables encountered in difficult piloting tasks in actual flight environments. The findings in the PT and 2T performance measures were reinforced on the bilinguals perceived effort, as shown in Figure 7.

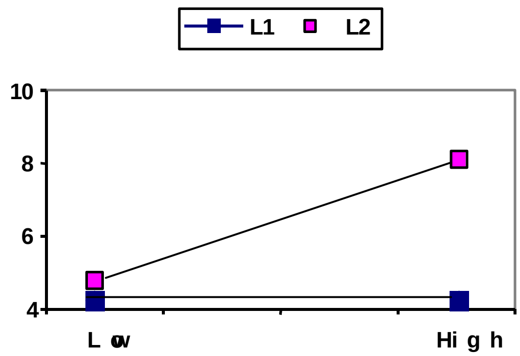


Figure 7. Perceived effort of balanced bilinguals at low to high difficulty levels for L1 and L2

CANADIAN IFR SIMULATION STUDY

Perhaps the strongest evidence to mislead one to question the notion that nonnative English speaking pilots incur greater workload in L2 airspace than their L1 peers is the bilingual air traffic communications simulation study (BICSS) conducted by the Canadian Ministry of Transport. It compared the safety of bilingual and unilingual air traffic control in the provinces of Quebec and Montreal. The findings of that study presented overwhelming evidence that bilingual air traffic control was safe (in Canada). With the exception of addressing a lost communication problem of simulated non-English speaking pilots who strayed into English-only airspace, BICSS did not address bilingual-pilot workload in L2 only airspace. It avoided the issue by citing existing procedural knowledge for lost communications--navigate first, then communicate--if you can. But without an operable transmitter and receiver, it would be impossible to measure communicative workload. The pilot need only to use his or her eyes, ears and piloting ability to attend to the navigational task Even with IFR flight, the pilot would fly his flight plan in event of communications failure.

For bilingual air traffic controllers, however, Transport Canada documents more communication errors in bilingual communication than in monolingual communications; however, the study found the differences insignificant and due to air traffic controller style rather than to the language factor. An example of the evidence was presented in Table 11. This report does not attempt to conduct a statistical analysis on the communications data because the communications data (Volume II of Transport Canada's report) was unavailable; rather, it focused on a discussion of the findings (Transport Canada, 1979)

"An unexpected benefit of the study was that some problems in non-language aspects of control procedures came to light. For example, it was discovered that many controllers paid insufficient attention to pilots' readbacks of clearances and infrequently caught pilot errors. A number of changes in air traffic controllers' Manual of Operations resulted from these findings" (Borins, 1983: 204). Due to this finding, it is proposed that the report measured only the overt performance of the bilingual controllers, or in Dornic's terms, the primary task. Had the controllers faithfully attended to their secondary task (attending to pilot errors and readbacks), more confidence could be placed in the findings and inferences on communicative workload of bilinguals in L2 airspace.

"The team established that the whole question of readbacks required a serious review by existing authorities. This review would of necessity, embrace the legal requirement for pilots to readback critical portions of all clearances and instructions as well as the requirement for controllers to listen to the pilot readbacks and correct errors. It would appear from results obtained, both in simulation and in real world reviews, that the frequency of controllers detecting incorrect readbacks is too low" (Transport of Canada, 1979: 50).

In the simulation study by Transport Canada (1979) Proulx and Deschenes minimized the difference in bilingual/unilingual readback errors, which was one of the controllers' secondary task performance measures, and they emphasized "the most important errors"--loss of separation, which was the bilingual controller's primary (PT) performance measure. Based on the dual task paradigm, and Kahneman's single resource theory (1973)--the supply and demand model of limited attentional resources--the secondary task performance measure is what is used to reveal controllers' hidden costs of maintaining primary task performance. By minimizing the importance of the 2T data, the potential to discover differences between the overt and covert (hidden) costs of bilingual and unilingual controllers' ability to maintain separation of aircraft was lost. The report, therefore, remains indecisive in terms of measuring communicative workload of bilinguals in L2 airspace.

Conclusion

Some might argue that Dornic's findings from simplified “real-life” L1 and L2 tasks conducted in the laboratory could not be generalized to L1 or L2 performance in dynamic flight. For example, it could be argued that context-based visual cues in actual fight enhance L1 and L2 performance by increasing pilots’ situational awareness (Endsley, 1988). Thus, context-based evaluations may be perceived as easier for the subject than laboratory-based evaluations. On the other hand, Dornic's laboratory experiments on noise (1978) could be perceived less difficult in terms of filtering load because the verbal component of noise in those laboratory experiments were not discernable. This removed the semantic processing component of workload.

The important point is that in actual flight requiring two way pilot-controller communications, pilots must filter meaningful from meaningless words and phrases. Furthermore, they must filter discernable word components of noise that apply to their specific call sign. Therefore, filtering load in complex airspace appears greater than filtering load in Dornic’s (1978) laboratory experiments.

Laboratory experiments had a major flaw--the experimenter's method of determining if a bilingual S was truly a "balanced bilingual." This concept was not standardized in the majority of the findings presented in this study, and this was a concern for external validity.

Applied to dynamic domains, the concept of standardization for selecting balanced bilingual Ss appeared increasingly important for measuring communication workload, and accentuated the expanded role of standardization as it related to bilingual Ss' level of piloting skill and language skill. Alessi (1988) demonstrated that visual noise had both positive and negative effects on workload depending on stage of training and experience level (novice or expert and so forth). Therefore, how well laboratory data on bilingual workload predicted real world bilingual pilot L1 and L2 workload appeared speculative. For reasons already mentioned, the findings of Transport Canada (1979) on bilingual pilot-controller workload in L2 airspace remain indecisive. A combination of the two approaches seems beneficial with technology.

Further research in simulation-based testing may discover more valid and practicable language measures with the advancement of technology (Alessi and Trollip, 1991: 232-235). But in the absence of more decisive and objective data, laboratory reports are considered relevant and valuable concerning communication tasks in dynamic domains.

Identifying potential communication problems may lead to discovery of cockpit design solutions. Improving on these communication problems with proper cockpit design to enhance communication efficiency may enhance pilots' and controllers' situational awareness. Future research should focus on developing objective assessment instruments that measure pilot communication workload to aid in the design of cockpits that reduce workload for nonnative speaking pilots in L2 airspace. This is especially critical in non-routine communications or emergency situations that depend on time sensitive reactions to verbal input.

REFERENCES

1. Albert, M. L. (1978). The bilingual brain. New York: Academic.
2. Baddeley, A. D. (1966). The capacity for generating information by randomization. Quarterly Journal of Experimental Psychology, 18, 119-129.
3. Billings, C. E., & Cheaney, E. S. (Eds.). (1981). Information transfer problems in the aviation system (NASA Technical paper No. 1875). Washington, DC: National and Space Administration. Scientific and Technical Information Branch.
4. Borg, G., Bratfisch, O., & Dornic, S. (1971). On the problems of perceived difficulty. Scandinavian Journal of Psychology, 12, 249-260.
5. Borins, S. F. (1983). The language of the skies: The bilingual air traffic control conflict in Canada. Kingston and Montreal: McGill-Queen's University.
6. Brown, I. D. (1978). Dual task methods for assessing workload. Ergonomics, 21(3), 221-224.
7. Brown, I. D. (1962). Measuring the spare mental capacity of car drivers by a subsidiary auditory task. Ergonomics, 5, 247-250.
8. Brown, I. D., & Poulton, E. C. (1961). Measuring the spare mental capacity of a car driver by a subsidiary task Ergonomics, 4, 35-40.
9. Cheng, L. (Ed.). (1995). Integrating language and learning for inclusion: An Asian/Pacific

focus (culture, rehabilitation and education). San Diego: Singular Publishing.
10. Charlton, S. G. (1996). Mental workload test and evaluation. In Thomas G. Obrien and Samuel G. Charleton (Eds.) (1996). Handbook of human factors testing and evaluation. New Jersey: Laurence Erlbaum.
11. Clark, H., & Schaffer, E. F. (1987). Collaborating on contributions to conversation. Language and Cognitive Processes, 2(1), 19-41.
12. Dornic, S. (1980a). Information processing and language dominance. International Review of Applied Psychology: the Journal of the International Association of Applied Psychology, 29, 119-140.
13. Dornic, S. (1980b). Language dominance, spare capacity and perceived effort in bilinguals. Ergonomics, 23(4), 369-377.
14. Dornic, S. (1979). Information processing in bilinguals. Psychological Research, 40, 329-348.
15. Dornic, S. (1978). Noise and language dominance (Rep. No. 536). Stockholm, Sweden: University of Stockholm, Department of Psychology.
16. Dornic, S. (1977a). Information processing and bilingualism (Rep. No. 510). Stockholm, Sweden: University of Stockholm, Department of Psychology.
17. Dornic, S. (1977b). Selective retention in bilingual tasks. In Dornic (Ed.), Attention and performance VI. (pp. 669-678.). Hilsdale, N.J.: Erlbaum.
18. Dornic, S. (1969). Verbal factor in number perception. Acta Psychologica, 29, 393-399.
19. Dupoux, E., & Mehler, J. (1992). Unifying awareness and on-line studies of speech: A tentative framework. In J. Alegria. D. Holender, J. Junca de Morias, & M. Radeau (Eds.), Analytical approaches to human cognition (pp.59-75). Paris: Elsevier Science Publishers B. V.
20. Endsley, M. R. (1988). Design and evaluation for situational awareness enhancement. Proceedings of the Human Factors Society, 32nd Annual Meeting, 97-101.
21. FAA (1988). Call to action: A joint FAA industry partnership to improve pilot/controller communications [Brochure]. Washington DC: Author
22. Fishman, J. & cooper, R. (1969). Alternative measures of bilingualism. Journal of Learning and Verbal Behavior, 8, 276-282.
23. Goguen, J., & Linde, C. (1983). Linguistic methodology for the analysis of aviation accidents (NASA Contractor Report 3741). Washington DC: NASA
24. Hamers, J. F., & Lambert, W. E. (1972). Bilingual interdependencies in auditory perception. Journal of Verbal Learning and Verbal Behavior, 11, 303-310.
25. Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. Human Factors, 26(5), 545-556.
26. Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice Hall.
27. Lambert, W. E. (1955). Measurement of the linguistic dominance of bilinguals. Journal of Abnormal and Social Psychology, 50, 197-200.
28. Linde, C. (1988). The quantitative study of communicative success: Politeness and accidents in aviation discourse. Language in Society, 17(3), 375-399.
29. Macnamara, J. (1967). The bilingual's linguistic performance: A psychological overview. Journal of Social Issues, 23(2), 58-77.
30. Michon, J. A. (1966). Tapping as a measure of perceptual motor load. Ergonomics, 9(5), 401-412.
31. Morrow, D., Rodvold, M. & Lee A. (1994). Non-routine transactions in controller-pilot communications. Discourse Processes, 17, 235-258.
32. Noble, C. E. (1997). The aviation English problem in America: Can a real-time based flight simulator help? [Available on the internet at http://www.aznet.net/~cnoble/aviationarticle] Annals, 1, (Isahaya City, Japan: Nagasaki Wesleyan Language Research Institute) 37-47.
33. Preston, M. S., & Lambert, W. E. (1969). Interlingual interference in a bilingual version of the Stroop Color Word Task. Journal of Verbal Learning and Verbal Behavior, 8, 295-301.
34. Ragan, P. H. (1997). Aviation English: an introduction. The Journal of Aviation/Aerospace Education and Research, 7(2), 25-36.
35. Robinson, C. P. and Eberts. (1987). Comparison of speech and pictorial displays in a cockpit environment. Human Factors, 29(1), 31-44.
36. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643-661.
37. Transport Canada (1979, January). Report of the bilingual IFR communications simulation studies: Vol. 1. Summary report and recommendations (Cat. No.: T52-53/1979-1). Canada: Minister of Supply and Services.
38. Vreuls, D. & Obermayer, R. W. (1985). Human system performance measurement in training simulators. Human Factors, 27(3), 241-250.

Author Note

Clifford E. Noble, Embry-Riddle Language Institute; Embry-Riddle Aeronautical University.

The paper is part of an ongoing review of the literature of the author's dissertation study at United States International University, San Diego CA, entitled Simulation-based Assessment as a Predictor of Nonnative English Student Pilots' Language Performance in American Airspace. Appreciation is extended to my committee chairperson, Dr. Sally Jensen, for her critical review, to Dr. John Wise of Embry-Riddle Aeronautical University's Human Factor's Engineering Department for his leadership, and finally to Dr. Pamela Tsang of Wright State University's Department of Experimental Psychology in Dayton Ohio, for her academic expertise and encouragement. The research reported does not reflect the views of either of the aforementioned universities.

Correspondence concerning this article should be addressed to Cliff Noble, who is now at Embry-Riddle Aeronautical University, Department of Humanities (ERLI), 600 S. Clyde Morris Blvd, Daytona Beach, Florida 32114. Electronic mail may be sent to [cliffnoble@aviationenglishtesting.com].

NOTES

[1]It is important to keep in mind that different workload metrics often produce different measurements of workload. Charleton (1996) discusses different workload metrics and their advantages and disadvantages. This report assumes that attention is the appropriate model for measuring workload for difficult tasks. The author couches the analysis of the subject experiments in Kahneman's undifferentiated single-resource theory (1973). The theory states that humans possess a limited supply of attentional resources. When demand for these resources exceeds supply, performance falters. When secondary tasks (2T) compete for available resources, the performance on 2T can indicate the "spare capacity" remaining to attend to anything else, providing that the primary task (PT) performance remains relatively stable. "Spare capacity" indexes workload. The author uses limited attentional resource theory and the concept of spare capacity to articulate measures of bilingual pilot workload in L1 and L2 airspace.

[2]See note in Table 2 for statistical method used.

[3]Similarities of stimulus will be discussed in more detail in the findings of experiment 2.

[4]A comprehensive study on the predictors of bilingual criteria are in Fishman and Cooper (1969).

[5] Note in table 5 that the combination of more difficult tasks were the control and interference tasks, which were labeled C1 through C6 and T1 through T6, respectively. The test for determining balanced bilingualism was gleaned from the lowest task-difficulty category (Basic). This is an important consideration for weighing validity of findings on applied studies on bilingualism. Caution should be used when making generalizations because many findings were based on the assumption that the Ss were balanced bilinguals. No standard definition for balanced bilinguals existed. This paper, however, limits itself to experiments of Lambert and two of his colleagues, and to Dornic; therefore there is a "within researcher" control of the term "balanced bilingual," which does not stray far from the mark.

[6] Reference was unavailable t for evaluating the statistical methodology of the experiment. The reference is : Dornic, S. (1978). The bilingual's performance: Language dominance, stress and individual differences. In Gerver, D. and Sinaiko, V. (eds.) Proceedings form the Interdiciplinary Symposium on Language, Interpretation, and Communication. New York: Plenum Press.

Table 1
Reaction Time Analysis of Non-native and Native French Bilinguals Majoring in French

Undergraduates				Graduates				Native French			
Ss	D	P	t*	Ss	D	P	t*	Ss	D	P	t*
U1	273	16	3.39	G1	-196	-11	2.08	F1	50	3	1.02
U2	172	8	2.33	G2	-86	-5	1.10	F2	-17	-1	.31
U3	235	10	1.73	G3	132	7	2.42	F3	-213	-9	1.84
U4	288	17	4.15	G4	207	8	1.95	F4	85	4	.97
U5	261	11	1.67	G5	212	10	3.14	F5	-335	-13	3.60
U6	129	7	2.61	G6	209	10	2.35	F6	-97	-5	1.19
U7	79	4	1.22	G7	178	9	2.37	F7	11	1	.21
U8	372	15	2.21	G8	93	6	1.49	F8	12	1	.16
U9	217	11	2.52	G9	39	2	.48	F9	-32	-2	.53
U10	169	10	1.72	G10	74	5	.94	F10	28	2	.35
U11	214	12	3.02	G11	250	14	3.12	F11	-204	-12	2.36
U12	282	15	3.28	G12	149	9	1.81	F12	-111	-4	.73
U13	354	20	3.86	G13	131	6	1.59	F13	-192	-9	3.02
U14	277	15	2.28	G14	19	1	.24	F14	-58	-3	.77

Note. Entries in column D are the absolute differences in time (hundredths of a second) taken to respond to 16 French and 16 English directions. A minus sign indicates that the responses to French directions were faster. Entries in column P are the absolute differences divided by the differences between the language speeds.

* Significance levels of t scores, one-tailed test, with 15 df: 1.75 $p > .05$; 2.60 when $p < .01$. From "Measurement of the Linguistic Dominance of Bilinguals," by Wallace E. Lambert, 1955, Journal of Abnormal and Social Psychology, 50, p. 198. Copyright 1955 by the American Psychological Association. Adapted and reprinted with permission.

Table 2
Mean Time Scores (sec) for English-Hungarian and English-French Bilinguals on Six Tasks

	English-Hungarian bilinguals (N=8)			English-French bilinguals (N=8)		
	English color-words	Hungarian color-words	Asterisks	English color-words	French color-words	Asterisks
English Response	1 111.5	2 103.9	3 76.6	1 99.2	2 94.2	3 67.5
Hungarian or French Response	4 98.3	5 113.1	6 77.1	4 102.2	5 100.1	6 68.5

Note. The six means were treated as one task, and the task effect was found to be significant for both groups using a one way analysis of variance model with repeated measures. $F(5,35) = 24.43$ $p<.01$ for English-Hungarian Ss and $F(5, 35) = 34.99$ $p<.01$ for English French Ss. The Neuman-Keuls multiple comparison procedure for the .05 level of significance was applied.

From "Interlingual Interference in a Bilingual Version of the Stroop color-Word Task," by Malcom S. Preston and Wallace E. Lambert, 1969, Journal of Verbal Learning and Verbal Behavior, 8, p. 296-297. Copyright 1969 by the American Psychological Association. Adapted and reprinted with permission.

Table 3

Mean Time Scores (sec) for Groups 1 and 2 on the Six Numbered Tasks

	Group 1 (N=8)			Group 2 (N=8)		
	English color-words	German color-words	Wavy lines	English color-words	German color-words	Wavy lines
English Response	1 100.8	2 101.6	3 65.5	1 94.8	2 82.3	3 67.6
German Response	4 105.2	5 99.7	6 66.6	4 85.6	5 93.9	6 66.0

Note. A three-way analysis of variance was conducted on the C Cards with the variables being word language, response language and Groups. Only significant interactions were the 3-way Group, word language, and response-language interaction, $F(1,14) = 23.67$, $p < .01$, and the Response Language/word Language interaction, $F(1, 14) = p < .05$.for exploration, the Neuman-Keuls procedure was applied to each of the four means within each group.

Table 4

Mean Time Scores (sec) for English-French Bilinguals Dominant in English on the Ten Numbered tasks

N = 8

	English color-words	French color-words	English noncolor-words	French noncolor-words	Letters
English Response	1 93.2	2 88.5	3 68.6	4 67.7	5 60.5
French Response	1 96.9	2 102.4	3 81.3	4 82.6	5 83.6

Note. A one-way analysis of variance was conducted and showed a significant task effect, $F(9, 63) = 34.57$, $p < .01$. The Newman-Keuls procedure was applied for the .05 level

Table 5

Basic, Control and Interference Tasks and Response Modes used to Measure Latency Effects of Pitch and Word on Auditory Perception of French and English Bilinguals

N = 16 French and English bilinguals

Response Modes	Basic (x100)	Control (x100)	Interference (x200)
Key Response	B1 (5, 6)	C1 (1:5, 2:6) C2 (3:5, 4:6)	T1 (1:5, 1:6, 2:5, 2:6) T2 (3:5, 3:6, 4:5, 4:6)
English Response	B2 (5, 6)	C3 (1:5, 2:6) C4 (3:5, 4:6)	T3 (1:5, 1:6, 2:5, 2:6) T4 (3:5, 3:6, 4:5, 4:6)
French Response	B3 (5, 6)	C5 (1:5, 2:6) C6 (3:5, 4:6)	T5 (1:5, 1:6, 2:5, 2:6) T6 (3:5, 3:6, 4:5, 4:6)

Note. 1 = low, 2 = high , 3 = basse, 4 = haute, 5 = 110 cps, 6 = 175 cps. Pitch = cps. Colons represented combinations of tasks conducted simultaneously. Commas annotated separate tasks.
From "Bilingual Interdependencies in Auditory Perception," by Malcom S. Preston and Wallace E. Lambert, 1969, Journal of Verbal Learning and Verbal Behavior, 11, p. 304-305. Table created form the text by Clifford E. Noble. Portions copyright 1972 by the American Psychological Association. Adapted and reprinted with permission.

Table 6

Mean Latencies for the 15 Tasks

	Stimuli		
Responses	Tones	English words	French words
Keypress	B1 = 664.1	C1 = 726.7 T1 = 910.9	C4 = 747.9 T4 = 872.4
English	B2 = 776.8	C2 = 904.1 T2 = 1115.1	C6 = 843.2 T6 = 1029.3
French	B3 = 762.6	C3 = 899.1 T3 = 1057.7	C5 = 933.9 T5 = 1084.6

Note. N = 16 French and English bilinguals
From "Bilingual Interdependencies in Auditory Perception," by Malcom S. Preston and Wallace E. Lambert, 1969, Journal of Verbal Learning and Verbal Behavior, 11, p. 306. Copyright 1972 by the American Psychological Association. Reprinted with permission.

Table 7

Mean Latencies for Various Stimulus Response Conditions

	Type of Task			
Type of stimulus	Monolingual SE + RE	Monolingual SF + RF	Bilingual SE + RF	Bilingual SF + RE
Congruent in control task (C)	904.2	933.9	899.1	843.2
Congruent in noncongruent task (TC)	1050.9	1010.6	1008.9	973.2
Incongruent in noncongruent task (TINC)	1179.4	1158.9	1106.6	1076.9

Note. N = 16 Ss. Entries are in msec.
From "Bilingual Interdependencies in Auditory Perception," by Malcom S. Preston and Wallace E. Lambert, 1969, Journal of Verbal Learning and Verbal Behavior, 11, p. 306. Copyright 1972 by the American Psychological Association. Reprinted with permission.

Table 8

Mean Latencies for Keypress Responses, Verbal Responses in Monolingual and Bilingual Tasks to B, C, TC and TINC

	Type of task		
		Verbal response	
Type of stimulus	Key press responses	Monolingual SE, RE and SF, RF	Bilingual SE, RF and SF,RE
Nonverbal (B)	664.1	973.2	769.7
Congruent in control task (C)	737.3	919.1	871.1
Congruent in noncongruent task (TC)	874.6	1031.4	991.1
Incongruent in noncongruent task (TINC)	960.4	1169.2	1091.3

Note. N = 16 Ss. Entries are in msec.
From "Bilingual Interdependencies in Auditory Perception," by Malcom S. Preston and Wallace E. Lambert, 1969, Journal of Verbal Learning and Verbal Behavior, 11, p. 307. Copyright 1972 by the American Psychological Association. Reprinted with permission.

Table 9

Mean Percentage Errors for Key Press and Verbal Responses in Monolingual and Bilingual Tasks to B, C, TC and TINC

	Type of task		
		Verbal response	
Type of stimulus	Key responses	Monolingual SE + RE and SF + RF	Bilingual SE + RF and SF + RE
Nonverbal (B)	0.3	1.2	
Congruent in control task (C)	1.1	1.4	1.2
Congruent in noncongruent task (TC)	4.2	6.3	2.9
Incongruent in noncongruent task (TINC)	20.7	32.8	25.6

Note. N = 16 Ss.

From "Bilingual Interdependencies in Auditory Perception," by Malcom S. Preston and Wallace E. Lambert, 1969, Journal of Verbal Learning and Verbal Behavior, 11, p. 308.

Table 10
Group Averages of Reported Number

Condition of Stimulation and language		Actual number 13	15	17
Flashes	Swedish	10.6	12.1	13.0
	English	9.4	10.2	11.0
Tones	Swedish	11.1	13.1	13.8
	English	9.9	11.0	10.8
Vibration	Swedish	11.2	12.7	13.8
	English	9.7	11.3	11.1

Note. N = 16 French and English bilinguals
From "Verbal Factor in Number Perception," by Stanislav Dornic, 1969, Acta Psychologica, p. 395. Copyright 1969 by North-Holland Publishing. Reprinted with permission (author deceased).

Table 11

Controller Language Use Bilingual and English Communication Errors for Phase I and Phase II

		Bilingual No. Errors	Rate/100	English No. Errors	Rate/100
James Bay					
Language use		15	0.68	0	
No. Transmissions		2219		2223	
Granby-Sherbrooke					
Language use		28	0.87	1	0.03
No. Transmissions		2219		2223	
Combined Enroute					
Language use		43	0.79	1	0.02
No. Transmissions		5421		5462	
Tones	Swedish		11.1	13.1	13.8
	English		9.9	11.0	10.8
Low Arrival					
Language use		24	0.38	0	
No. Transmissions		6340		6540	
Arrival Sequencer					
Language use		49	0.69	0	
No. Transmissions		7131		7158	
Tones	Swedish	11.1	13.1	13.8	
	English	9.9	11.0	10.8	
Departure					
Language use		48	0.75	0	
No. Transmissions		6376		6449	
Positions Combined					
Language use		121	0.74	0	
No. Transmissions		19847		20147	
Enroute and Terminal Combined					
Language use		164	0.65	1	
No. Transmissions		2219		2223	

Note. For total errors, $p < .05$. Language use includes initial use of incorrect language and correction within a transmission or a correction in language following a failure to obtain a reply. Modified form Appendix J of Ministry of Transport (1979) Report of the Bilingual IFR Communications Simulations Study, Volume 1. Reprinted with permission.

FIGURES

Figure 1. Fatal misunderstanding of automated cockpit warning in English to bilingual Chinese pilots to "Pull-up" while shooting an Instrument Landing System (ILS) approach.

Figure 2. Average number of stimuli of flashes, vibrations and auditory senses counted by self-proclaimed balanced English and Swedish bilinguals in a covert counting task compared to actual number presented.

Figure 3. Response time and performance on random L1 and L2 decoding and detection workload tasks with verbal stimulus.

Figure 4. Response time and performance on random L1 and L2 decoding and detection workload tasks with verbal stimulus.

Figure 5. Primary task performance at low to high difficulty levels for L1 and L2.

Figure 6. Secondary task performance at low to high difficulty levels for L1 and L2.

Figure 7. Perceived effort of balanced bilinguals at low to high difficulty levels for L1 and L2.